



TITLE:

# Multi-armed banditsとある非協力ゲームの関連について(確率ゲーム理論とその周辺)

AUTHOR(S):

YOSHIDA, Yuji

---

CITATION:

YOSHIDA, Yuji. Multi-armed banditsとある非協力ゲームの関連について(確率ゲーム理論とその周辺). 数理解析研究所講究録 1991, 741: 15-23

ISSUE DATE:

1991-01

URL:

<http://hdl.handle.net/2433/102126>

RIGHT:

# Mutli-armed bandits と ある非協力ゲームの関連について

千葉大学教養部 吉田祐治 (Yuji YOSHIDA)

## 1. d-armed Markov bandit processes.

$d$  : number of arms ( positive integer ).

$N = \{ 0,1,2,\dots \}$  : time space.

$(\Omega^i, \mathcal{F}^i, P^i)$  : probability spaces (  $i = 1, \dots, d$  ).

$X^i = (X_t^i, \mathcal{F}_t^i, P^i)_{t \in N}$  : mutually independent Markov chains

with Borel state spaces  $E^i$  (  $i = 1, \dots, d$  ).

$\{\mathcal{F}_t^i\}_{t \in N}$  is an increasing family of completed sub- $\sigma$ -fields of  $\mathcal{F}^i$

$X = (X_s)_{s \in T} = (X_{s1}^1, \dots, X_{sd}^d)_{s \in T} = (s^1, \dots, s^d) \in T$  : d-parameter process with state space  $E$

$T = N^d$  : time space.  $E = \prod_{i=1}^d E^i$  : state space.  $\Omega = \prod_{i=1}^d \Omega^i$  : path space.

$P = \prod_{i=1}^d P^i$  : probability measure.  $\mathcal{F}_s = \mathcal{F}_{s1}^1 \otimes \dots \otimes \mathcal{F}_{sd}^d$  for  $s = (s^1, \dots, s^d) \in T$ .

$T(\in N)$  : terminal time.

$N(e, T) = \{ \text{even } t : 0 \leq t < T \}$ ,  $N(o, T) = \{ \text{odd } t : 0 \leq t < T \}$ , for  $T(\in N)$

$\beta$  : discount rate (  $0 < \beta < 1$  ).  $\mathbf{0} = (0, \dots, 0) \in T$ .  $e_i = (0, \dots, 0, 1, 0, \dots, 0) \in T$ .

$f^i, g^i$  : fixed bounded measurable function on  $E^i$ .

$h, k$  : fixed bounded measurable function on  $E$ .

## 2. Strategies.

Strategies when player A moves first and second does player B ( first-type strategies

$$\pi = \{ \pi(t) \}_{t \in \mathbb{N}} = \{ (\pi^1(t), \dots, \pi^d(t)) \}_{t \in \mathbb{N}} \text{ and}$$

$$\sigma = \{ \sigma(t) \}_{t \in \mathbb{N}} = \{ (\sigma^1(t), \dots, \sigma^d(t)) \}_{t \in \mathbb{N}}$$

are  $\mathbf{T}$ - valued stochastic sequences on  $(\Omega, \mathcal{F})$  satisfying the following ( i ), ( ii ) and ( iii )

i)  $\pi(0) = \sigma(0) = \mathbf{0}$ .

ii) For all even  $t \in \mathbb{N}$  it holds that

$$\pi(t+1) = \sigma(t) + e_i \quad \text{for some } i = 1, \dots, d \quad \text{and}$$

$$\sigma(t+1) = \sigma(t).$$

For all odd  $t \in \mathbb{N}$  it holds that

$$\sigma(t+1) = \pi(t) + e_i \quad \text{for some } i = 1, \dots, d \quad \text{and}$$

$$\pi(t+1) = \pi(t).$$

iii) For all  $t \in \mathbb{N}$  and all  $s \in \mathbf{T}$  it holds that  $\{ \pi(t) = s \} \in \mathcal{F}_s$  and  $\{ \sigma(t) = s \} \in \mathcal{F}_s$ .

$$S(\mathbf{F}) = \{ \text{all first-type strategies } (\pi, \sigma) \text{ starting from } \mathbf{0} \},$$

$$MS(\mathbf{F}) = \{ \text{all Markov strategies } (\pi, \sigma) (\in S(\mathbf{F})) \},$$

$$MS(\mathbf{F}; 1) = \{ \text{all first-type one-step Markov strategies } \pi \}.$$

Strategies when player B moves first and second does player A ( second-type ).

$$\pi = \{ \pi(t) \}_{t \in \mathbb{N}} \text{ and } \sigma = \{ \sigma(t) \}_{t \in \mathbb{N}}$$

are  $\mathbf{T}$ - valued stochastic sequences on  $(\Omega, \mathcal{F})$  satisfying the following ( i ), ( ii ) and ( iii )

( i )  $\pi(0) = \sigma(0) = \mathbf{0}$ .

( ii ) For all even  $t \in \mathbb{N}$  it holds that

$$\sigma(t+1) = \pi(t) + e_i \quad \text{for some } i = 1, \dots, d \quad \text{and}$$

$$\pi(t+1) = \pi(t).$$

For all odd  $t \in \mathbb{N}$  it holds that

$$\pi(t+1) = \sigma(t) + e_i \quad \text{for some } i = 1, \dots, d \quad \text{and}$$

$$\sigma(t+1) = \sigma(t).$$

( iii ) For all  $t \in \mathbb{N}$  and all  $s \in \mathbb{T}$  it holds that  $\{ \pi(t) = s \} \in \mathcal{F}_s$  and  $\{ \sigma(t) = s \} \in \mathcal{F}_s$ .

$S(S) = \{ \text{all second-type strategies } (\pi, \sigma) \text{ starting from } \mathbf{0} \},$

$MS(S) = \{ \text{all Markov strategies } (\pi, \sigma) ( \in S(S) ) \}$  and

$MS(S; 1) = \{ \text{all second-type one-step Markov strategies } \sigma \}.$

$$D(F; \sigma) = \{ \pi : (\pi, \sigma) \in S(F) \}, \quad D(F; \pi) = \{ \sigma : (\pi, \sigma) \in S(F) \}.$$

$$D(S; \pi) = \{ \sigma : (\pi, \sigma) \in S(S) \} \text{ and } D(S; \sigma) = \{ \pi : (\pi, \sigma) \in S(S) \}.$$

### 3. Expected rewards in bandit games.

For  $(\pi, \sigma) ( \in S(F) )$ , player A's expected values from an initial state  $x ( \in E )$  to a terminal time  $T$  are defined by

$$\begin{aligned} R_{A,F,T}^{\pi,\sigma}(x) = & \mathbf{E}^x \left[ \sum_{i=1}^d f^i( X_{\pi i(1)}^i ) ( \pi^i(1) - \sigma^i(0) ) \right. \\ & \left. + \beta^2 \sum_{i=1}^d f^i( X_{\pi i(3)}^i ) ( \pi^i(3) - \sigma^i(2) ) + \dots \right] \end{aligned}$$

$$+ \beta^{T-2} \sum_{i=1}^d f^i( X_{\pi^i(T-1)} ) ( \pi^i(T-1) - \sigma^i(T-2) ) + \beta^T h( X_{\sigma(T)} ) ] \quad \text{when } T \text{ is even,}$$

and

$$\begin{aligned} R_{A,F,T}^{\pi,\sigma}(x) = & E^x \left[ \sum_{i=1}^d f^i( X_{\pi^i(1)} ) ( \pi^i(1) - \sigma^i(0) ) \right. \\ & + \beta^2 \sum_{i=1}^d f^i( X_{\pi^i(3)} ) ( \pi^i(3) - \sigma^i(2) ) + \dots \\ & \left. + \beta^{T-1} \sum_{i=1}^d f^i( X_{\pi^i(T)} ) ( \pi^i(T) - \sigma^i(T-1) ) + \beta^T h( X_{\pi(T)} ) \right] \quad \text{when } T \text{ is odd.} \end{aligned}$$

For short

$$(1) \quad R_{A,F,T}^{\pi,\sigma}(x) = E^x \left[ \sum_{t \in \mathbf{N}(e,T)} \beta^t \sum_{i=1}^d f^i( X_{\pi^i(t+1)} ) ( \pi^i(t+1) - \sigma^i(t) ) \right. \\ \left. + \beta^T h( X_{\max\{\pi(T), \sigma(T)\}} ) \right],$$

Then player B's expected values are

$$(2) \quad R_{B,F,T}^{\pi,\sigma}(x) = E^x \left[ \sum_{t \in \mathbf{N}(o,T)} \beta^t \sum_{i=1}^d g^i( X_{\sigma^i(t+1)} ) ( \sigma^i(t+1) - \pi^i(t) ) \right. \\ \left. + \beta^T k( X_{\max\{\pi(T), \sigma(T)\}} ) \right],$$

Hence we put

$$(3) \quad R_{A,F,T}^{\pi^*,\sigma}(x) = \sup_{\pi \in D(F;\sigma)} R_{A,F,T}^{\pi,\sigma}(x) \quad \text{for } x \in E, \text{ and}$$

$$(4) \quad R_{B,F,T}^{\pi,*}(x) = \sup_{\sigma \in D(F;\pi)} R_{B,F,T}^{\pi,\sigma}(x) \quad \text{for } x \in E.$$

Then we shall call the following game when player A moves first first-type bandit games abbreviated as FBG) : For each T, to find strategies  $(\pi^*, \sigma^*) \in S(F)$  such that

$$R_{A,F,T}^{\pi^*,\sigma^*} = R_{A,F,T}^{*,\sigma^*} \text{ and } R_{B,F,T}^{\pi^*,\sigma^*} = R_{B,F,T}^{\pi^*,*}.$$

For a second-type strategy  $(\pi, \sigma) \in S(S)$ , player A's expected values are defined b

$$(5) \quad R_{A,S,T}^{\pi,\sigma}(x) = E^x \left[ \sum_{t \in \mathbf{N}(o,T)} \beta^t \sum_{i=1}^d f^i( X_{\pi^i(t+1)} ) ( \pi^i(t+1) - \sigma^i(t) ) \right]$$

$$+ \beta^T h( X_{\max\{\pi(T), \sigma(T)\}} ) ] .$$

Then player B's expected values are

$$(6) \quad R_{B,S,T}^{\pi,\sigma}(x) = E^x \left[ \sum_{t \in N(e,T)} \beta^t \sum_{i=1}^d g^i( X_{\sigma^i}^i(t+1) ) ( \sigma^i(t+1) - \pi^i(t) ) + \beta^T k( X_{\max\{\pi(T), \sigma(T)\}} ) \right] .$$

Hence we put

$$(7) \quad R_{A,S,T}^{\pi^*,\sigma}(x) = \sup_{\pi \in D(S; \sigma)} R_{A,S,T}^{\pi,\sigma}(x) \quad \text{for } x \in E, \text{ and}$$

$$(8) \quad R_{B,S,T}^{\pi,*}(x) = \sup_{\sigma \in D(S; \pi)} R_{B,S,T}^{\pi,\sigma}(x) \quad \text{for } x \in E.$$

Then we shall call the following game when player A moves second second-type ba games ( abbreviated as SBG ) : For each T, to find strategies  $(\pi^*, \sigma^*) \in S(S)$  such th

$$R_{A,S,T}^{\pi^*,\sigma^*} = R_{A,S,T}^{\pi^*,\sigma^*} \text{ and } R_{B,S,T}^{\pi^*,\sigma^*} = R_{B,S,T}^{\pi^*,*}.$$

When the terminal time  $T = 0$ , for convenience we define

$$(9) \quad R_{A,F,0}^{\pi,\sigma} = h \text{ and } R_{B,F,0}^{\pi,\sigma} = k \text{ for all } (\pi, \sigma) \in S(F), \text{ and}$$

$$(10) \quad R_{A,S,0}^{\pi,\sigma} = h \text{ and } R_{B,S,0}^{\pi,\sigma} = k \text{ for all } (\pi, \sigma) \in S(S).$$

#### LEMMA 1.

For strategies  $(\pi, \sigma) \in S(F)$  and  $(\pi', \sigma') \in S(S)$ , there exist Markov strategies  $\pi_M$ ;  $0, \sigma$ ,  $\sigma_M \in D(F; 0, \pi)$ ,  $\pi'_M \in D(S; 0, \sigma')$  and  $\sigma'_M \in D(S; 0, \pi')$  such that

$$R_{A,F,T}^{\pi_M, \sigma} = R_{A,F,T}^{*, \sigma}, \quad R_{B,F,T}^{\pi, \sigma_M} = R_{B,F,T}^{\pi, *},$$

$$R_{A,S,T}^{\pi'_M, \sigma'} = R_{A,S,T}^{*, \sigma'} \text{ and } R_{B,S,T}^{\pi', \sigma'_M} = R_{B,S,T}^{\pi', *}.$$

#### 4. A value iteration and optimal strategies.

##### ITERATION 1.

##### Subroutine ( A ) :

( A. 0 ) Put  $U_{A,F,0} = U_{A,S,0} = h$ .

( A. F. 1 ) Put  $U_{A,F,1}(x) = \max_{1 \leq i \leq d} E^x \left[ f^i(X_1^i) + \beta U_{A,S,0}(x^1, \dots, X_1^i, \dots, x^d) \right]$

$x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\pi_1^* \in MS(F; 1)$ .

( A. S. 1 ) Put  $U_{A,S,1}(x) = E^x \left[ \beta U_{A,F,0}(X_{\sigma_1^*}(1)) \right]$  for  $x \in E$

th  $\sigma_1^* \in MS(S; 1)$  given by ( B. S. 1 ).

( A. F. 2 ) Put  $U_{A,F,2}(x) = \max_{1 \leq i \leq d} E^x \left[ f^i(X_1^i) + \beta U_{A,S,1}(x^1, \dots, X_1^i, \dots, x^d) \right]$

$x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\pi_2^* \in MS(F; 1)$ .

( A. S. 2 ) Put  $U_{A,S,2}(x) = E^x \left[ \beta U_{A,F,1}(X_{\sigma_2^*}(1)) \right]$  for  $x \in E$

th  $\sigma_2^* \in MS(S; 1)$  given by ( B. S. 2 ).

.....

( A. F. r+1 ) Put  $U_{A,F,r+1}(x) = \max_{1 \leq i \leq d} E^x \left[ f^i(X_1^i) + \beta U_{A,S,r}(x^1, \dots, X_1^i, \dots, x^d) \right]$

$x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\pi_{r+1}^* \in MS(F; 1)$ .

( A. S. r+1 ) Put  $U_{A,S,r+1}(x) = E^x \left[ \beta U_{A,F,r}(X_{\sigma_{r+1}^*}(1)) \right]$  for  $x \in E$

th  $\sigma_{r+1}^* \in MS(S; 1)$  given by ( B. S. r+1 ).

.....

( A. F. T ) Put  $U_{A,F,T}(x) = \max_{1 \leq i \leq d} E^x \left[ f^i(X_1^i) + \beta U_{A,S,T-1}(x^1, \dots, X_1^i, \dots, x^d) \right]$

$x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\pi_T^* \in MS(F; 1)$ .

( A. S. T ) Put  $U_{A,S,T}(x) = E^x \left[ \beta U_{A,F,T-1}(X_{\sigma_T^*(1)}) \right]$  for  $x \in E$

with  $\sigma_T^* \in MS(S; 1)$  given by ( B. S. T ).

**Subroutine ( B ) :**

( B. 0 ) Put  $U_{B,F,0} = U_{B,S,0} = k$ .

( B. F. 1 ) Put  $U_{B,F,1}(x) = E^x \left[ \beta U_{B,S,0}(X_{\pi_1^*(1)}) \right]$  for  $x \in E$

with  $\pi_1^* \in MS(F; 1)$  given by ( A. F. 1 ).

( B. S. 1 ) Put  $U_{B,S,1}(x) = \max_{1 \leq i \leq d} E^x \left[ g^i(X_1^i) + \beta U_{B,F,0}(x^1, \dots, X_1^i, \dots, x^d) \right]$

for  $x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\sigma_1^* \in MS(S; 1)$ .

( B. F. 2 ) Put  $U_{B,F,2}(x) = E^x \left[ \beta U_{B,S,1}(X_{\pi_2^*(1)}) \right]$  for  $x \in E$

with  $\pi_2^* \in MS(F; 1)$  given by ( A. F. 2 ).

( B. S. 2 ) Put  $U_{B,S,2}(x) = \max_{1 \leq i \leq d} E^x \left[ g^i(X_1^i) + \beta U_{B,F,1}(x^1, \dots, X_1^i, \dots, x^d) \right]$

for  $x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\sigma_2^* \in MS(S; 1)$ .

.....

( B. F. r+1 ) Put  $U_{B,F,r+1}(x) = E^x \left[ \beta U_{B,S,r}(X_{\pi_{r+1}^*(1)}) \right]$  for  $x \in E$

with  $\pi_{r+1}^* \in MS(F; 1)$  given by ( A. F. r+1 ).

( B. S. r+1 ) Put  $U_{B,S,r+1}(x) = \max_{1 \leq i \leq d} E^x \left[ g^i(X_1^i) + \beta U_{B,F,r}(x^1, \dots, X_1^i, \dots, x^d) \right]$

for  $x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\sigma_{r+1}^* \in MS(S; 1)$ .

.....

( B. F. T ) Put  $U_{B,F,T}(x) = E^x \left[ \beta U_{B,S,T-1}(X_{\pi_T^*(1)}) \right]$  for  $x \in E$

with  $\pi_T^* \in MS(F; 1)$  given by ( A. F. T ).



$$(B, S, T) \text{ Put } U_{B,S,T}(x) = \max_{1 \leq i \leq d} E^x \left[ g^i(X_1^i) + \beta U_{B,F,T-1}(x^1, \dots, X_1^i, \dots, x^d) \right]$$

for  $x = (x^1, \dots, x^d) \in E$ . We define a Markov strategy  $\sigma_T^* \in MS(S; 1)$ .

We define strategies  $(\pi^*, \sigma^*) \in MS(F; T)$  and  $(\pi'^*, \sigma'^*) \in MS(S; T)$  by

- 1)  $(\pi^*, \sigma^*) = [\pi_T^*, \sigma_{T-1}^*, \pi_{T-2}^*, \sigma_{T-3}^*, \dots, \pi_4^*, \sigma_3^*, \pi_2^*, \sigma_1^*]$  for even  $T$ ,
- 2)  $(\pi^*, \sigma^*) = [\pi_T^*, \sigma_{T-1}^*, \pi_{T-2}^*, \sigma_{T-3}^*, \dots, \sigma_4^*, \pi_3^*, \sigma_2^*, \pi_1^*]$  for odd  $T$ ,
- 3)  $(\pi'^*, \sigma'^*) = [\sigma_T^*, \pi_{T-1}^*, \sigma_{T-2}^*, \pi_{T-3}^*, \dots, \sigma_4^*, \pi_3^*, \sigma_2^*, \pi_1^*]$  for even  $T$ , and
- 4)  $(\pi'^*, \sigma'^*) = [\sigma_T^*, \pi_{T-1}^*, \sigma_{T-2}^*, \pi_{T-3}^*, \dots, \pi_4^*, \sigma_3^*, \pi_2^*, \sigma_1^*]$  for odd  $T$ .

#### THEOREM 1.

$(\pi^*, \sigma^*)$  is an optimal strategy for FBG, and  $(\pi'^*, \sigma'^*)$  is an optimal strategy for SBG. Moreover  $U_{A,F,T}$  and  $U_{B,F,T}$  are each player's optimal values for FBG and  $U_{A,S,T}$  and  $U_{B,S,T}$  are each player's optimal values for SBG :

- (i)  $U_{A,F,T} = R_{A,F,T}^{\pi^*, \sigma^*} \geq R_{A,F,T}^{\pi, \sigma^*}$  for every  $\pi \in D(F; \sigma^*)$ .
- (ii)  $U_{B,F,T} = R_{B,F,T}^{\pi^*, \sigma^*} \geq R_{B,F,T}^{\pi^*, \sigma}$  for every  $\sigma \in D(F; \pi^*)$ .
- (iii)  $U_{A,S,T} = R_{A,S,T}^{\pi'^*, \sigma'^*} \geq R_{A,S,T}^{\pi, \sigma'^*}$  for every  $\pi \in D(S; \sigma'^*)$ .
- (iv)  $U_{B,S,T} = R_{B,S,T}^{\pi'^*, \sigma'^*} \geq R_{B,S,T}^{\pi'^*, \sigma}$  for every  $\sigma \in D(S; \pi'^*)$ .

For Markov strategies  $\pi \in MS(F; 1)$  and  $\sigma \in MS(S; 1)$  we shall introduce the following semi-linear operators  $S^\pi$  and  $S^\sigma$  on the space of all bounded measurable functions on  $E$

$$(15) \quad S^\pi \phi(x) = E^x \left[ \sum_{i=1}^d f^i(X_{\pi^i(1)}^i) \pi^i(1) + \beta \phi(X_{\pi(1)}) \right] \text{ for } x \in E, \text{ and}$$

$$(16) \quad S^\sigma \phi(x) = E^x \left[ \sum_{i=1}^d g^i(X_{\sigma^i(1)}^i) \sigma^i(1) + \beta \phi(X_{\sigma(1)}) \right] \text{ for } x \in E$$

for bounded measurable functions  $\phi$  on  $E$ . Then

### COROLLARY 1.

For  $r = 0, \dots, T-1$  (i) ~ (iv) hold :

$$(i) \quad U_{A,F,r+1}(x) = \max_{1 \leq i \leq d} E^x \left[ f^i(X_1^i) + \beta U_{A,S,r}(x^1, \dots, X_1^i, \dots, x^d) \right]$$

for  $x = (x^1, \dots, x^d) \in E$ .

$$(ii) \quad U_{A,S,r+1}(x) = E^x \left[ \beta U_{A,F,r}(X_{\sigma_{r+1}^*(1)}) \right] \text{ for } x \in E,$$

where  $\sigma_{r+1}^* \in MS(S; 1)$  given by (iv).

$$(iii) \quad U_{B,F,r+1}(x) = E^x \left[ \beta U_{B,S,r}(X_{\pi_{r+1}^*(1)}) \right] \text{ for } x \in E,$$

where  $\pi_{r+1}^* \in MS(F; 1)$  given by (i).

$$(iv) \quad U_{B,S,r+1}(x) = \max_{1 \leq i \leq d} E^x \left[ g^i(X_1^i) + \beta U_{B,F,r}(x^1, \dots, X_1^i, \dots, x^d) \right]$$

for  $x = (x^1, \dots, x^d) \in E$ .

$$(v) \quad U_{A,F,r+1} = S^{\pi_{r+1}^*} U_{A,S,r}, \quad U_{A,S,r+1} = \beta P^{\sigma_{r+1}^*} U_{A,F,r},$$

$$U_{B,F,r+1} = \beta P^{\pi_{r+1}^*} U_{B,S,r}, \quad U_{B,S,r+1} = S^{\sigma_{r+1}^*} U_{B,F,r}.$$